DOCUMENT RESUME

ED 328 568 TM 016 059

AUTHOR Veccia, Ellen M.; Schroeder, David H.
TITLE Construction and Analysis of Semantic Properties of an Alternate Form of the Personality Worksample. Technical Report 1990-6.
INSTITUTION Johnson O'Connor Research Foundation, New York, NY. Human Engineering Lab.
PUB DATE Dec 90
NOTE 44p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Adults; Correlation; *Personality Measures; *Psychological Testing; *Semantics; *Test Construction; Test Format; Test Items; Test Validity
IDENTIFIERS *Objectivity; *Subjectivity

ABSTRACT

A set of 150 experimental personality items was constructed for an alternate form of the word association personality worksample developed by the Johnson O'Connor Research Foundation. The items were intended to possess several semantic properties hypothesized to facilitate discrimination between objective and subjective examinees. Specifically, the items were intended to be items that were contrasts, were concrete, had a physical or sensational reference, and were capable of generating alternative responses other than the contrast response. The items were also expected to elicit the keyed response from at least 15% of the examinees. Each of the experimental items was administered to approximately 100 examinees; a total of 1,541 examinees participated in the study. The importance of each of the subject property to an item's performance was examined by having three raters indicate the degree to which each stimulus-keyed response word pair possessed each of the properties. The mean ratings were then correlated with the items' correlations with total scores on Worksample 35 A (referred to as item-total correlations). Interactions among the ratings were also analyzed. Correlations revealed that the contrast property was the best predictor of item-total correlation with Worksample 35 A scores. A semantic analysis of Worksample 35 A items by M. Daniel, and administration instructions for the worksample are appended. Six data tables and one figure are included. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED328568

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

ROBERT KYLE

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CONSTRUCTION AND ANALYSIS OF SEMANTIC PROPERTIES OF AN ALTERNATE FORM OF THE PERSONALITY WORKSAMPLE

Ellen M. Veccia

and

David H. Schroeder

JOHNSON O'CONNOR RESEARCH FOUNDATION, INC.

Technical Report 1990-6

December 1990

TM016059

COPYRIGHT © 1990 BY JOHNSON O'CONNOR RESEARCH FOUNDATION, INCORPORATED
ALL RIGHTS RESERVED

Construction and Analysis of Semantic Properties of an Alternate Form of the Personality Worksample

Ellen M. Veccia and David H. Schroeder

ABSTRACT

A set of 150 experimental Personality items was constructed for an alternate form of the Foundation's word association Personality worksample. The items were intended to possess several semantic properties hypothesized to facilitate discrimination between objective and subjective examinees. Specifically, the items were intended to be items that were contrasts, were concrete, had a physical or sensational reference, and were capable of generating alternative responses other than the contrast response. The items were also expected to elicit the keyed response from at least 15% of the examinees to whom they were presented. Each of the experimental items was administered to about 100 examinees.

The importance of each of these properties to an item's performance was examined by having three raters indicate the degree to which each stimulus-keyed response word pair possessed each of the properties. The mean ratings were then correlated with the items' correlations with total scores on Worksample 35 A (referred to as item-total correlations). Interactions between the ratings were also analyzed.

The correlations revealed that the contrast property was the best predictor of item-total correlation with 35 A scores. Significant correlations between the other properties and for all interaction terms were dissipated when the effect of the contrast property was removed by means of partial correlations. Previous hypotheses regarding the characteristics of effective items, as well a potential explanation for the previous failures to generate an acceptable alternate form of the Personality worksample were discussed in light of these new data.

Of the items examined here, 50 were selected as expected scored items for an alternate form of the Personality worksample. These 50 contrast items were significantly correlated with Worksample 35 A scores ($r > .30$) and had keyed responses that were given by at least 15% of the examinees. The alternate form

was constructed such that ten warm-up (non-scored) items are administered first, and then the 50 expected scored contrast items are alternated with noncontrast filler items, which are included to deter examinees from detecting a pattern of contrast items. A small number of the filler items, however, are similar to coordinates and will be explored as potential scored items when large-sample data have been collected.

CONTENTS

	Page
Introduction	1
Method	4
Item Construction	4
Raters	5
Subjective Ratings	5
Examinees	6
Administration	8
Results	8
Subjective Ratings	8
Construction of an Alternate Form of the Personality Worksample	21
Discussion	23
References	25
Appendix A	26
Appendix B	33

LIST OF TABLES AND FIGURE

Page

Table 1 Descriptive Sample Statistics 7

Table 2 Item Analysis and Average Ratings for the Experimental Personality Items 10

Table 3 Correlations Between Item Properties and Item-Total Correlation 17

Figure 1 Distribution of Correlations With 35 A Scores as a Function of Mean Contrast Ratings 18

FROM APPENDIX A

Table 1 Ten Best Items of Worksample 35 AI 28

Table 2 Next Ten Best Items of Worksample 35 AI 29

Table 3 Categorization of Items from Worksample 35 AI 30

ACKNOWLEDGMENTS

We gratefully acknowledge the staffs of the Atlanta, Boston, Chicago, Dallas/Ft. Worth, Denver, Houston, Los Angeles, New Orleans, New York, San Francisco, Seattle, Tampa, and Washington, D.C., testing offices for their assistance in item construction and data collection.

The Research Department staff entered the data and provided valuable assistance in ensuring the accuracy of the data presented in this report.

Cassandra Swartz was largely responsible for organizing and analyzing these data. Her contribution to this project greatly facilitated its progress.

INTRODUCTION

The Foundation's Personality worksample is a word association task. A word association task requires an individual to respond to a stimulus word with the first word that comes to mind. The pioneer empirical studies using the word association task were conducted by Kent and Rosanoff (1910) and Wells (1911), who attempted to use this task to classify psychiatric patients. Although these early classification attempts were not successful, they yielded two useful lists of stimulus words together with the most common responses to those words. The most common response is the response most frequently given for each stimulus word.

In the mid-1920s Johnson O'Connor used words from the Kent-Rosanoff and Wells lists to develop a word association task designed to measure a different psychological dimension than had been studied by earlier investigators (O'Connor, 1928; see also O'Connor (1934) for a detailed description of this early work). More specifically, O'Connor used the word association task as a tool for differentiating personality types.

According to O'Connor, there are two alternative personality types: objective and subjective. The majority of people have objective personalities. Objective persons attend more to the external world than to their own personal sensations and perceptions, while subjective individuals attend to their own personal sensations and perceptions more strongly than they attend to the external world (O'Connor, 1928).

These rather different perspectives on the world are indeed reflected by performance on the word association task. When the test was scored by counting the number of most common responses given by the examinee, the results showed that one group of examinees obtained higher scores than did a second, smaller group of examinees. That is, objective individuals had a higher number of most common responses than did subjective individuals (O'Connor, 1934). An alternative way of expressing the typical patterns of scores is that subjective individuals had a larger number of personalized responses (and, therefore, fewer common responses) than did objective individuals.

During the course of revising early versions of the Personality worksample, it became apparent that some of the stimulus/common-response word pairs were more effective than others in differentiating between objective and subjective individuals. In order to describe the characteristics of these word pairs, O'Connor (1934) developed the following semantic classification scheme:

1. "The class *Opposites* consists of adjectives and adverbs to which the most common response is the direct opposite: Black-*white*, Slow-*fast*, High-*low* . . . .

2. "The type *Coordinates* contains no adjectives or adverbs, but in some instances stimulus and response are as clearly opposite as in the previous class, such as, Health-*sickness*, Follow-*lead*, Death-*life*. In other cases, stimulus and response, while belonging to the same class, oppose one another in that they are mutually exclusive. Man-*woman*, Salt-*pepper*, Lightning-*thunder* are not opposites in the same sense as White-*black*, but are complementary . . . . In this class probably belong such reactions as Trouble-*pleasure*, and Scatter-*gather* . . . .

3. "The class synonyms explains itself, Wish-*desire*.

4. "In the 4th type, namely *Classifications*, the response names the class to which the stimulus belongs: thus Butterfly-*insect* . . . .

5. *"Associations* differ from coordinates with which they are most apt to be confused in that stimulus and response do not belong to the same type or category. Umbrella-*rain*, although associated, are not of the same category.

"To the word TABLE *chair* is the coordinate, *furniture* the classification, *food* the association. There is no commonly given opposite or synonym.

6. "The class *Miscellaneous* contains some responses of distinct types, such as Fruit-*apple*, the response a specific example of the stimulus word, examples of which occur too seldom to warrant the creation of a class. In other replies, such as Seed-*plant*, the response may be either noun or verb, and the response changes class accordingly. Other replies are classified as miscellaneous because of the difficulty of knowing the subject's intent . . . ." (O'Connor, 1934, pp. 215-216).

Opposites (adjectives and adverbs) and Coordinates together form the class *Contrasts*. The class Coordinates includes both opposing words that are neither adjectives nor adverbs (e.g., Follow-*lead*) and complementary words (e.g., Man-*woman*).

Application of this semantic classification to early forms of the Personality worksample revealed that the stimulus/most-common-response word pairs classified as *Contrasts* were the most effective items in differentiating the two personality types (O'Connor, 1934). This semantic analysis also revealed the

necessity to examine characteristics other than commonality of the response. For some stimulus words, the most common response was given more frequently by subjective individuals than objective individuals, a trend that is opposite in direction to the typical pattern. Thus, in terms of item effectiveness in discriminating objective and subjective individuals, O'Connor concluded that: "Commonality of response is not the critical factor. Type of response stands out predominantly" (O'Connor, 1934, p. 216).

The importance of the type of response led to a more in-depth examination of the second, third, and fourth most common responses. Here again, an important pattern of results was revealed. For some stimulus words, especially those of the contrast type, these less common responses correlated strongly with the most common responses. More importantly, when these other responses were included in the scoring of the test, the reliability of the test was increased. Since scoring responses that were correlated with the most common responses contributed positively to the test, the scoring method was changed from counting the number of most common responses to counting most common responses and responses that were correlated with most common responses. The new scoring method assigned a score of 1 to "significant" responses and a score of 0 to other responses. Significant responses were identified on the basis of correlation with the total score.

To summarize, the semantic classification described above resulted in a revised scoring method for the Personality worksample because it provided new information regarding the most useful responses for differentiating between objective and subjective individuals. In addition, it was expected that the semantic classification scheme would be useful in developing alternate forms of this worksample because it would allow researchers to generate items that would be efficient in discriminating between objective and subjective examinees.

In spite of the understanding added by the classification findings, there is a substantial history of unsuccessful attempts to develop an alternate form of the Personality worksample (for details see the *Personality Worksample 35 A Manual*). The alternate forms typically possessed high internal reliability but did not correlate sufficiently highly with the standard form, Form A, of the worksample (Daniel, 1977; *Personality Worksample 35 A Manual*). Several hypotheses have been offered regarding the failure to develop an acceptable alternate form, including the possibility that important characteristics of effective stimulus words were not adequately represented in alternate forms (Daniel, 1977).

In an unpublished semantic analysis of Worksample 35 AI (Daniel, 1977, reprinted in Appendix A of this report), Daniel explored this idea. His analysis suggested that the stimulus words on Worksample 35 A that best discriminated between objective and subjective individuals, besides being contrasts, tended to

relate to the body (e.g., Hand-*foot*) or to physical sensations (e.g., Soft-*hard*). Abstract words and concrete nouns that were contrasts were the next best items, and abstract and concrete items that were not contrasts were the worst items. Thus, effective 35 A items appeared to be contrasts but also to relate to the body or to physical sensations. Daniel speculated that previously developed alternate forms may have failed because the stimulus words were too abstract, were unlikely to elicit responses other than the obvious opposite (e.g., In-*out*), and were particularly unlikely to stimulate the personal images thought to be characteristic of subjective persons.

The study presented in this report was designed to evaluate empirically the properties of effective Personality items suggested by Daniel's analysis. This was accomplished by having persons rate each stimulus-response word pair in the present study on the degree to which it possessed these properties. Items that corresponded most strongly to the characteristics prescribed by Daniel were expected to show the highest correlations with total scores on Worksample 35 A. Thus, items that were rated as contrasts, having physical or sensational referents, and having the potential to elicit responses other than the obvious opposite were expected to demonstrate the highest correlations with total scores on Worksample 35. In addition, since the best items in Daniel's analysis were also concrete (or at least nonabstract), we also evaluated item concreteness in relation to item effectiveness.

A secondary goal of this research was to develop an alternate form of the Personality worksample that would correlate acceptably with Worksample 35 A.

METHOD

Item Construction

Daniel's unpublished item analysis (Daniel, 1977; see Appendix A) was used to provide guidelines for the construction of potential items. As detailed by Schroeder (Statistical Bulletin 1988-3), the 13 testing offices were provided with the semantic and descriptive guidelines of good items, based on Daniel's findings. These guidelines described good items as possessing four characteristics: contrast items, concrete items, items that refer to the body or sensations, and items that have a potential to generate alternative responses. In addition, the contrast response for each item was expected to be a relatively common response (i.e., the response was expected to be given by at least 10 to 15% of the examinees tested). Each testing office was asked to generate 20 potential stimulus-response pairs according to these guidelines. Schroeder, in concert with the given testing office, then selected 10 items for each office. If an identical item had already been

assigned to a testing office, then the next best item was selected. The Chicago and New York offices each were assigned two sets of experimental items. (Seattle also administered Chicago's second set of items.) There were a total of 150 experimental items. In addition, ten noncontrast items were selected to serve as "filler" items on the experimental forms, to break up the pattern of contrast items. Thus, there were a total of 15 sets of 10 experimental items each, which were combined with the 10 filler items to yield 15 forms of 20 items each. Although the testing offices did not refer to previous forms of the Personality worksample in writing their items, many of the experimental items replicated items on earlier forms.

Raters

Subjective ratings (see below) were obtained from three raters (two females and one male). Two of the raters had substantial experience with the Personality worksample and had knowledge of its history. The third rater was comparatively naive regarding Worksample 35 A.

Subjective Ratings

The set of 150 experimental items, together with the 10 filler items, were placed on a single list in order to obtain subjective ratings of the characteristics (see earlier description). These characteristics were expected to affect the efficiency of each item in discriminating between objective and subjective individuals, and therefore, to determine how effective the item would be in eliciting responses that would correlate with Worksample 35 A. For some of the experimental items, there was more than one contrast response that was expected to correlate with 35 A. In such instances, a separate rating was obtained for each of the contrast responses. Inclusion of all contrast responses resulted in a set of 200 stimulus-response pairs for the rating task. The first 10 of these pairs were the filler items.

The raters were instructed to indicate, on 5-point scales, the degree to which each stimulus-response pair possessed the four characteristics noted earlier: (a) Contrast, where 5 = definitely a contrast and 1 = definitely not a contrast; (b) Non-abstractness, where 5 = definitely *not* abstract (concrete) and 1 = definitely abstract; (c) Physical reference (reference to body parts, physical actions, or physical sensations/perceptions), where 5 = definitely refers to one of the listed referents and 1 = definitely does not refer to one of the physical referents; and (d) Alternative responses (capacity to evoke personalized responses in addition to the contrast response), where 5 = has very good potential to evoke alternative (personalized) responses and 1 = has poor potential. The high point of each scale (5) indicated that the item possessed the property of interest, and the low point (1) indicated that the item lacked this property. (Although the raters were instructed to rate items for nonabstractness, we will, for simplicity, refer to this property as concreteness throughout the remainder of this report.)

Raters were also requested to provide comments that they expected to be helpful in determining the appropriateness of each stimulus-response pair for an alternate form of the Personality worksample. Raters completed the rating forms at their convenience and returned the forms to the Research Department for analysis.

When more than two raters provide judgments of the same stimulus or event, reliability across the raters is determined by calculating the *effective reliability* (Rosenthal, 1982). The effective reliability is analogous to calculating the projected reliability for a test with a larger number of items than the test being studied (the Spearman-Brown formula), except that effective reliability is based on the mean reliability coefficient obtained from the total number of pairwise comparisons. Thus, the *r* for effective reliability is somewhat different from that used to calculate projected reliability. Here, *r* refers to the mean of the coefficients that were obtained by calculating all the pairwise interrater correlations. The total possible number of pairwise comparisons is determined by Formula 1.

Formula (1): $$Total\ pairwise\ comparisons = \frac{n(n-1)}{2} \quad ,$$

Formula 2 is the formula for determining the effective reliability across raters.

Formula (2): $$R = \frac{n\bar{r}}{1 + (n - 1)\bar{r}} \quad ,$$

where R = the effective reliability, $\bar{r}$ = the mean reliability obtained from pairwise comparisons, and n = the number of raters. Notice that when there are only two raters, Formula 1 is equal to one, thus, the effective reliability is equal to the actual reliability between the two raters.

Examinees

The examinees were paying clients (code "00") unselected for sex and age--that is, scholarship and public information examinees were not included. A total of 1,541 examinees participated in the study, with approximately 100 examinees tested on each of the experimental forms. With respect to the total sample, examinees' ages ranged from 14 to 67 years, with a mean (*M*) = 28.48 and standard deviation (*SD*) = 9.95. Sex was nearly equally represented, with males representing 51.6% of the sample and females 48.4%. Table 1 provides a detailed description of the sample for each testing office, including examinees' average raw

score for the standard Personality form, Worksample 35 A. Examinees were generally white and characterized economically as middle to upper-middle class. The examinees tested by the Foundation typically have at least some college education or are college-bound. Given that the examinees were paying clients of the Foundation, they were self-selected for aptitude testing, usually for the purpose of career guidance.

Table 1

Descriptive Sample Statistics

Testing office	Sample size	% females	% males	Age range	Median age	Mean raw score, 35 A
Atlanta	106	50.9[a]	48.1[a]	14-61	21	16.7
Boston	101	51.5	48.5	15-54	21	16.5
Chicago I	100	52.0	48.0	16-52	25	17.9
Chicago II	94	37.2	62.8	14-56	27	16.6
Dallas	100	47.0	53.0	15-52	19	16.1
Denver	101	49.5	50.5	15-60	27	15.3
Houston	101	39.6	60.4	15-60	19	17.5
Los Angeles	102	43.1	56.9	14-57	22	15.5
New Orleans	112	44.6	55.4	15-64	22	14.1
New York I	102	59.8	40.2	16-56	23	15.6
New York II	100	51.0[a]	48.0[a]	15-57	22	14.1
San Francisco	106	47.2	52.8	14-59	25	16.0
Seattle	107	47.7	52.3	15-67	27	14.4
Tampa	104	50.0	50.0	16-54	27	17.2
Washington, DC	105	53.3	46.7	15-59	19	17.0

Note. Total N = 1,541.
[a]Sex was missing for one case.

Administration

The sets of experimental items were distributed to the testing offices for administration. Each testing office was provided with testing forms that contained the ten experimental items assigned to that office, with these items separated from each other by the ten filler items. Experimental and filler items were alternated so that examinees would be less inclined to fall into a pattern of responding to each stimulus word with a contrast type of response (Statistical Bulletin 1988-3).

The experimental forms were administered to examinees immediately upon completion of the Personality Worksample 35 A, but prior to the repeating of any 35 A items on which the examinee blocked. Test administrators followed the same procedures in administering the experimental form as those used for Worksample 35 A. The administration instructions for Worksample 35 A are presented in Appendix B. Briefly, the items are presented orally at the rate of about one item every 2 seconds. Items on which examinees block (i.e., fail to provide a response within their normal response time) are recorded (see Appendix B). Since the experimental form was given immediately after Worksample 35 A, examinees responded to items on the experimental form based on the instructions for 35 A. From the examinee's perspective, the experimental form was not a distinct test from the standard Personality worksample. The number of examinees tested on each group of experimental items corresponds to the sample size for each testing office shown in Table 1. The experimental forms were administered in the latter half of 1987 except for the second Chicago form, which was administered in early 1988.

RESULTS

Subjective Ratings

The first step in analyzing the rating results was to determine how well the raters agreed in their ratings of the item characteristics by determining the effective reliability for each of the properties that were rated. The results of this analysis indicated that the raters agreed most strongly for the contrast and physical reference ratings. For the contrast ratings, the effective reliability (R) was .85, and for the physical reference ratings, $R = .84$. The concreteness ratings yielded an effective reliability that was somewhat lower, $R = .65$.

For the alternative response ratings, the rater having the least familiarity with the Personality worksample had substantially different ratings than did the other two raters. Consequently, these ratings were not included in the analysis, and the reliability for this rating was determined by the reliability coefficient r (Pearson) for

Raters 1 and 2. The reliability for the alternative response rating was r = .35. The low coefficient for the alternative response scale indicates that the raters often were not in agreement in their judgments of this property. These ratings were retained for the remainder of the analyses, but the results involving this scale must be interpreted with caution because the judgments were not reliable across raters. Different raters might yield somewhat different results. Also, the correlations between the ratings will be attenuated for their less-than-perfect reliabilities.

To assess how well the subjective ratings predicted item property, the mean ratings for the contrast, concreteness, and physical reference dimensions were calculated across the three raters. For the alternative response dimension the mean was calculated for Raters 1 and 2 only. The mean ratings for each stimulus and keyed response are shown in Table 2. The mean for all the contrast ratings was 3.78 (SD = 1.12); for the concreteness ratings, M = 4.15 (SD = .70); for the physical reference ratings, M = 2.98 (SD = 1.42); and for the alternative response ratings, the mean was 3.44 (SD = .95).

Since 5-point scales were used for the rating task, the mid-point of the rating scales was 3.0, halfway between 1 and 5. Thus, the means obtained here indicate that overall, the experimental items written by the testing offices possessed the desired properties and represented items that were concrete contrast items that had the potential to evoke personalized or alternative responses. There was, however, a moderate degree of variation among the word pairs on each of the dimensions as indicated by the SD for each scale. The physical reference property appears to be less common than the other properties among the experimental items--the mean rating for this scale was just below the midpoint of the scale and lower than the means of the other rating scales.

The last two columns of Table 2 show the proportion of examinees who responded to the stimulus with the expected keyed response (p) and the item-total correlation for each of the 200 word pairs. In general, the proportion of individuals who responded to the stimulus by giving the keyed response was in the expected range (at least 10 to 15%), with a mean p-value of .31 (SD = .25). There were, however, a few stimuli for which the expected keyed response was not given by any of the examinees (p = .00). As will be discussed, some of these items were ultimately dropped as potential test items, and others were retained, but a new keyed response was selected for them based on the analysis of the other responses given.

Item-total correlations indicate how well an item's score agrees with the total test score. In the present context, the item-total correlation indicates how well the item's scores correlate with scores on the established Worksample 35 A. The mean item-total correlation across the 200 items shown in Table 2 was .25 (SD = .16). The item-total correlations ranged from -.18 to .62. The mean item-total

Table 2

Item Analysis and Average Ratings for the Experimental Personality Items

Item number	Stimulus	Keyed response	Mean contrast[a]	Mean non-abstract-ness[a]	Mean physical reference[a]	Mean alternative response[b]	Proportion keyed response	Item-total correlation[c]
1.	Carpet	rug	1.67	5.00	1.33	3.50	.23	.20*
2.	Picture	frame	1.33	4.67	1.67	4.00	.37	.17*
3.	Spider	web	2.00	5.00	1.33	4.00	.44	.09*
4.	Pumpkin	orange	1.33	5.00	1.00	3.00	.16	-.05
5.	Jet	plane	1.00	5.00	1.00	2.50	.40	.19*
6.	Shoe	sock	2.67	5.00	3.67	4.00	.11	.13*
7.	Glitter	shine	1.00	3.67	3.00	4.00	.12	.08*
8.	Lock	key	2.33	4.67	1.67	2.50	.45	.08*
9.	Atom	molecule	2.67	3.33	1.00	3.00	.16	.12*
10.	Tape	recorder	1.33	4.67	1.00	2.00	.16	.02
11.	Straight	bent	5.00	4.00	2.67	4.00	.03	.13
12.	Straight	crooked	5.00	3.67	2.67	4.00	.20	.27*
13.	Elbow	knee	3.33	5.00	5.00	1.50	.15	.23
14.	Tall	short	5.00	4.33	4.33	4.50	.64	.40*
15.	Eye	ear	3.00	5.00	5.00	4.00	.05	.13
16.	Coat	hat	2.33	4.67	3.33	2.50	.06	.07
17.	Animal	vegetable	3.33	4.67	1.67	4.00	.03	-.01
18.	Animal	plant	3.33	4.67	1.67	4.50	.00	--
19.	Red	white	1.67	4.33	3.67	4.50	.18	.10
20.	Red	blue	1.67	4.33	3.67	4.50	.19	.32*
21.	Red	green	2.67	4.33	3.67	4.50	.04	.14
22.	Laugh	cry	5.00	4.00	5.00	2.50	.35	.39*
23.	In	out	5.00	2.33	1.33	1.00	.77	.40*
24.	Left	right	5.00	2.67	2.00	1.50	.85	.25*
25.	Work	play	4.00	3.33	3.33	4.00	.43	.32*
26.	Hot	cold	5.00	4.67	5.00	3.50	.63	.49*
27.	Hot	cool	3.33	4.33	5.00	3.50	.03	.12
28.	Come	go	5.00	2.67	3.00	1.50	.63	.39*
29.	Under	over	5.00	2.67	2.00	1.00	.52	.44*
30.	Hide	seek	4.33	4.00	3.00	2.00	.40	.31*

Note. Items 1 to 10 were "filler" items. For them, $N = 1{,}541$. For Items 11 to 200, $94 \leq N \leq 107$.

[a]These values represent the mean rating across three raters.

[b]These values represent the mean rating across two raters.

[c]The item-total correlation here refers to the correlation between the item and the total score on Worksample 35 A.

*$\alpha < .01$.

Table continues

Table 2 (cont'd)

Item Analysis and Average Ratings for the Experimental Personality Items

Item number	Stimulus	Keyed response	Mean contrast[a]	Mean non-abstract-ness[a]	Mean physical reference[a]	Mean alternative response[b]	Proportion keyed response	Item-total correlation[c]
31.	Sing	dance	3.00	4.67	5.00	4.50	.09	.21
32.	Floor	ceiling	4.33	4.67	1.33	3.00	.26	.36*
33.	Curly	straight	4.67	4.33	3.67	4.00	.43	.44*
34.	Top	bottom	5.00	4.00	2.00	2.50	.57	.62*
35.	Day	night	5.00	4.33	2.33	3.00	.79	.44*
36.	Shout	whisper	4.33	4.67	5.00	4.00	.04	.23
37.	Aunt	uncle	4.67	4.67	1.00	4.00	.26	.31*
38.	Moist	dry	4.33	4.67	4.67	4.50	.17	.27*
39.	Love	hate	5.00	2.67	3.67	4.50	.56	.33*
40.	Read	write	3.67	3.33	3.67	3.00	.23	.23
41.	Warm	cold	3.67	4.33	5.00	4.00	.46	.45*
42.	Heaven	hell	5.00	3.67	1.33	2.50	.50	.38*
43.	Heaven	earth	3.33	4.00	1.33	3.50	.12	.16
44.	Paper	pencil	3.33	5.00	1.00	3.50	.23	.03
45.	Paper	pen	3.33	5.00	1.00	3.50	.13	.16
46.	Easy	hard	4.67	3.00	2.00	4.00	.69	.50*
47.	Army	Navy	3.67	4.33	1.00	2.00	.40	.27*
48.	Strong	weak	5.00	4.00	3.67	3.50	.46	.47*
49.	Throw	catch	4.33	4.67	5.00	4.00	.32	.24
50.	Down	up	5.00	3.00	1.33	2.00	.71	.51*
51.	Happy	sad	5.00	4.00	3.67	3.00	.63	.53*
52.	Sit	stand	4.00	4.33	4.33	4.50	.46	.36*
53.	Coffee	tea	3.33	5.00	2.33	3.50	.10	.15
54.	Cold	hot	5.00	4.67	5.00	5.00	.38	.37*
55.	Large	small	5.00	4.00	2.33	4.50	.50	.34*
56.	Stop	go	5.00	4.33	2.33	3.00	.71	.49*
57.	Stop	start	4.00	4.00	2.00	3.00	.04	.03
58.	Pencil	paper	3.33	5.00	1.33	2.00	.15	.04
59.	Pencil	pen	3.33	5.00	1.33	2.00	.31	.29*
60.	Fat	thin	5.00	4.67	4.00	5.00	.39	.38*

Note. Items 1 to 10 were "filler" items. For them, N = 1,541. For Items 11 to 200, 94 $\leq N \leq$ 107.

[a]These values represent the mean rating across three raters.

[b]These values represent the mean rating across two raters.

[c]The item-total correlation here refers to the correlation between the item and the total score on Worksample 35 A.

*$\alpha < .01$.

Table continues

Table 2 (cont'd)

Item Analysis and Average Ratings for the Experimental Personality Items

Item number	Stimulus	Keyed response	Mean contrast[a]	Mean non-abstract-ness[a]	Mean physical reference[a]	Mean alternative response[b]	Proportion keyed response	Item-total correlation[c]
61.	Pen	pencil	3.67	4.67	1.33	3.50	.45	.50*
62.	Pen	paper	3.00	4.67	1.33	3.50	.06	-.18
63.	Walk	run	3.67	4.67	5.00	3.50	.48	.38*
64.	Open	close	5.00	4.33	3.33	4.00	.40	.19
65.	Open	closed	4.00	4.00	2.00	4.00	.32	.26*
66.	Push	pull	4.67	4.33	4.67	1.50	.20	.12
67.	Mother	father	4.67	5.00	1.67	4.50	.65	.33*
68.	Mother	child	2.33	5.00	1.67	4.50	.02	.08
69.	Winter	summer	4.00	4.33	2.00	5.00	.35	.46*
70.	Question	answer	4.33	3.33	1.00	2.00	.71	.10
71.	Blonde	brunette	4.00	4.33	4.67	4.50	.45	.36*
72.	Short	long	4.67	4.33	3.00	4.50	.10	.07
73.	Short	tall	4.33	4.33	4.33	4.50	.59	.25
74.	Cow	pig	1.33	5.00	1.00	3.00	.05	-.08
75.	Cow	horse	1.33	5.00	1.00	3.00	.14	.32*
76.	Eat	drink	4.00	4.67	5.00	4.00	.13	.29*
77.	Hair	head	2.00	5.00	5.00	5.00	.04	.06
78.	Air	water	2.67	4.67	1.67	4.00	.08	.23
79.	Tight	loose	5.00	4.00	2.67	4.00	.42	.51*
80.	Give	receive	4.00	3.67	2.00	3.00	.20	-.12
81.	Give	take	4.67	4.00	2.00	3.00	.46	.38*
82.	Ear	eye	2.33	5.00	5.00	3.50	.01	-.07
83.	Ear	nose	2.00	5.00	5.00	3.50	.08	.27*
84.	Run	walk	4.00	4.67	5.00	3.00	.25	.38*
85.	Melt	freeze	4.00	4.33	3.33	2.50	.08	.29*
86.	Hungry	full	3.33	4.00	4.33	4.50	.04	.26*
87.	Hungry	thirsty	4.00	4.67	4.67	4.50	.13	.20
88.	Dull	sharp	5.00	4.00	2.33	4.00	.19	.27*
89.	Dull	bright	4.00	4.00	2.67	4.00	.04	.03
90.	Take	give	4.00	3.00	2.33	3.50	.40	.21

Note. Items 1 to 10 were "filler" items. For them, N = 1,541. For Items 11 to 200, $94 \leq N \leq 107$.

[a]These values represent the mean rating across three raters.

[b]These values represent the mean rating across two raters.

[c]The item-total correlation here refers to the correlation between the item and the total score on Worksample 35 A.

*$\alpha < .01$.

Table continues

Table 2 (cont'd)

Item Analysis and Average Ratings for the Experimental Personality Items

Item number	Stimulus	Keyed response	Mean contrast[a]	Mean non-abstract-ness[a]	Mean physical reference[a]	Mean alternative response[b]	Proportion keyed response	Item-total correlation[c]
91.	Catch	throw	4.00	4.67	4.67	4.00	.49	.23
92.	Loud	soft	4.33	4.00	4.00	4.00	.45	.43*
93.	Knee	elbow	3.00	5.00	5.00	3.00	.13	.12
94.	Back	front	4.33	4.00	2.67	2.50	.32	.21
95.	Rip	tear	1.33	4.33	3.67	3.50	.70	.18
96.	Eyes	ears	3.33	5.00	5.00	3.00	.08	.10
97.	Eyes	nose	2.67	5.00	5.00	3.00	.22	.27*
98.	Circle	square	4.33	4.67	1.00	2.50	.63	.29*
99.	Over	under	5.00	3.00	1.33	1.00	.67	.22
100.	Front	back	4.67	4.00	2.67	3.00	.84	.43*
101.	North	south	5.00	4.00	1.67	3.00	.90	.39*
102.	Big	small	4.33	4.33	3.33	4.50	.44	.27*
103.	Big	little	5.00	4.00	3.67	4.50	.26	.21
104.	First	last	5.00	3.67	1.33	2.00	.65	.36*
105.	Inside	outside	5.00	4.00	2.00	2.00	.69	.50*
106.	Bread	butter	4.00	5.00	3.00	3.50	.34	.23
107.	Narrow	wide	5.00	4.00	2.67	4.00	.47	.48*
108.	Early	late	5.00	3.33	1.00	3.00	.71	.46*
109.	Rise	fall	4.33	3.67	2.67	2.50	.40	.27*
110.	Bow	tie	2.00	5.00	2.67	2.50	.17	-.00
111.	Bow	arrow	3.33	5.00	2.00	2.50	.48	.18
112.	Wet	dry	5.00	4.67	5.00	4.00	.59	.58*
113.	Light	dark	4.67	4.33	4.33	4.50	.30	.36*
114.	Light	heavy	4.00	4.33	4.33	4.50	.01	.02
115.	Arm	leg	3.67	5.00	5.00	3.50	.50	.36*
116.	Cool	warm	4.33	4.33	4.67	4.50	.05	.18
117.	Sharp	dull	5.00	4.33	3.00	4.50	.16	.40*
118.	Full	empty	4.67	4.33	2.67	4.00	.47	.45*
119.	Full	hungry	3.00	4.00	4.67	4.00	.00	--
120.	Night	day	5.00	4.67	1.67	3.50	.51	.43*

Note. Items 1 to 10 were "filler" items. For them, N = 1,541. For Items 11 to 200, $94 \leq N \leq 107$.

[a]These values represent the mean rating across three raters.

[b]These values represent the mean rating across two raters.

[c]The item-total correlation refers to the correlation between the item and the total score on Worksample 35 A.

*$a < .01$.

Table continues

Table 2 (cont'd)

Item Analysis and Average Ratings for the Experimental Personality Items

Item number	Stimulus	Keyed response	Mean contrast[a]	Mean non-abstract-ness[a]	Mean physical reference[a]	Mean alternative response[b]	Proportion keyed response	Item-total correlation[c]
121.	Up	down	5.00	3.33	1.67	1.00	.78	.51*
122.	Tall	short	4.67	4.00	4.33	4.50	.47	.43*
123.	Head	foot	3.00	5.00	4.33	3.00	.04	.21
124.	Head	toe	3.00	5.00	5.00	3.00	.06	.13
125.	Wrist	ankle	3.33	5.00	5.00	3.50	.06	.30*
126.	Sad	happy	5.00	3.33	4.00	3.50	.48	.58*
127.	Sweet	sour	5.00	4.33	5.00	4.50	.36	.42*
128.	Bumpy	smooth	3.67	4.33	4.33	4.50	.27	.51*
129.	Bumpy	flat	3.33	4.67	4.33	4.50	.00	--
130.	State	city	1.67	3.67	1.67	4.50	.06	.13
131.	Enter	exit	4.67	4.00	3.33	2.00	.48	.34*
132.	Leg	arm	3.00	5.00	5.00	3.50	.35	--[e]
133.	Taste	touch	2.33	4.67	5.00	4.50	.01	.03
134.	Taste	smell	3.33	4.67	5.00	4.50	.14	.25*
135.	Pleasant	unpleasant	5.00	3.33	2.00	4.00	.12	.26*
136.	Sight	sound	3.33	4.33	4.67	3.50	.04	.04
137.	Finger	toe	4.00	5.00	5.00	4.00	.05	.12
138.	Child	adult	3.00	4.67	1.00	3.50	.04	.35*
139.	Child	parent	3.33	4.67	1.00	3.50	.01	.06
140.	Tender	tough	3.67	4.00	4.00	4.50	.02	.13
141.	Home	away	1.33	4.00	1.00	5.00	.05	-.09
142.	Angry	happy	3.33	3.67	3.67	4.00	.04	.09
143.	Heart	head	1.67	4.67	5.00	4.00	.00	--
144.	Sick	well	4.67	4.00	4.33	4.50	.11	.25
145.	Sick	healthy	4.33	4.00	4.33	4.50	.12	.27*
146.	Fuzzy	smooth	3.33	4.00	4.33	4.00	.01	.18
147.	Pain	pleasure	4.33	4.00	4.33	3.50	.00	--
148.	Street	avenue	1.00	4.67	1.33	3.50	.06	-.02
149.	Wide	narrow	5.00	4.00	2.33	4.00	.37	.39*
150.	Above	below	5.00	3.33	1.67	1.00	.51	.57*

Note. Items 1 to 10 were "filler" items. For them, N = 1,541. For Items 11 to 200, $94 \leq N \leq 107$.

[a]These values represent the mean rating across three raters.

[b]These values represent the mean rating across two raters.

[c]The item-total correlation here refers to the correlation between the item and the total score on Worksample 35 A.

[e]Item-total correlation could not be calculated due to a keying error.

*$\alpha < .01$.

Table continues

Table 2 (cont'd)

Item Analysis and Average Ratings for the Experimental Personality Items

Item number	Stimulus	Keyed response	Mean contrast[a]	Mean non-abstract-ness[a]	Mean physical reference[a]	Mean alternative response[b]	Proportion keyed response	Item-total correlation[c]
151.	Spool	thread	3.00	4.33	1.67	2.00	.92	.24
152.	Bright	dim	4.33	4.00	4.00	4.50	.07	.26*
153.	Bright	dull	4.00	3.67	3.33	4.50	.04	.16
154.	Bad	good	4.67	3.00	2.00	2.50	.64	.56*
155.	Little	big	5.00	4.00	3.33	3.50	.48	.44*
156.	Right	wrong	4.67	2.67	1.00	2.00	.45	.27*
157.	Right	left	4.67	2.67	3.00	2.00	.15	.09
158.	Apple	orange	1.67	5.00	1.67	4.00	.21	.25
159.	Poor	rich	4.33	3.00	1.00	3.50	.56	.55*
160.	East	west	5.00	4.00	1.67	3.00	.80	.45*
161.	Spoon	fork	3.00	5.00	1.67	3.50	.72	.30*
162.	Handsome	ugly	4.33	3.67	4.00	4.00	.15	.28*
163.	Lift	drop	3.00	4.00	4.33	3.50	.05	.24
164.	Point	out	1.33	3.00	3.00	3.00	.00	--
165.	Hit	hard	1.33	4.67	4.33	3.50	.06	-.04
166.	Flip	flop	4.33	4.33	2.67	2.00	.13	.13
167.	Slip	slide	3.33	4.67	4.33	2.50	.15	.09
168.	See	hear	2.67	4.33	5.00	3.00	.06	.22
169.	Wash	dry	3.33	4.67	4.33	3.00	.16	.13
170.	Look	listen	3.00	4.33	4.67	3.00	.00	--
171.	Dry	wet	4.67	4.33	5.00	4.00	.70	.44*
172.	Dry	moist	3.67	4.00	5.00	4.00	.00	--
173.	Beard	mustache	2.00	5.00	5.00	4.00	.19	.24
174.	Old	young	4.67	4.00	3.33	4.00	.28	.01
175.	Old	new	4.33	3.33	2.33	4.00	.43	.41*
176.	Wrong	right	4.67	3.00	1.33	1.50	.86	.38*
177.	Major	minor	4.33	2.00	1.00	3.00	.50	.19
178.	Beginning	ending	4.67	3.00	1.00	3.00	.13	.27*
179.	Beginning	end	4.33	2.67	1.00	3.00	.75	.02
180.	Seek	find	4.00	3.33	2.33	3.50	.75	.03

Note. Items 1 to 10 were "filler" items. For them, $N = 1,541$. For Items 11 to 200, $94 \leq N \leq 107$.

[a]These values represent the mean rating across three raters.

[b]These values represent the mean rating across two raters.

[c]The item-total correlation here refers to the correlation between the item and the total score on Worksample 35 A.

*$\alpha < .01$.

Table continues

Table 2 (cont'd)

Item Analysis and Average Ratings for the Experimental Personality Items

Item number	Stimulus	Keyed response	Mean contrast[a]	Mean non-abstract-ness[a]	Mean physical reference[a]	Mean alternative response[b]	Proportion keyed response	Item-total correlation[c]
181.	Crooked	straight	4.00	3.33	2.33	4.50	.69	.25
182.	Friend	enemy	4.00	4.33	1.33	5.00	.28	.39*
183.	Friend	foe	4.33	4.00	1.33	5.00	.28	.18
184.	Good	bad	4.67	3.00	2.00	3.00	.72	.11
185.	Good	evil	4.33	3.33	2.00	3.00	.13	.16
186.	Boring	exciting	4.00	3.00	2.00	3.00	.11	.23
187.	Boring	interesting	3.67	3.00	2.00	3.00	.02	.08
188.	Fast	slow	5.00	4.00	3.33	3.50	.67	.59*
189.	Sail	boat	1.33	4.33	2.33	3.00	.50	.13
190.	Smart	dumb	4.33	3.00	2.67	4.00	.45	.41*
191.	Smart	slow	3.00	3.00	2.67	4.00	.01	.07
192.	Smart	stupid	4.33	3.33	2.67	4.00	.08	.14
193.	Funny	sad	3.00	3.33	3.00	4.00	.14	.31*
194.	Cloudy	sunny	3.67	5.00	4.00	3.50	.18	.23
195.	Cloudy	clear	3.67	5.00	4.00	3.50	.04	.11
196.	Inner	outer	5.00	3.00	1.67	1.00	.52	.39*
197.	Sun	moon	3.00	5.00	1.67	2.50	.31	.25
198.	Son	daughter	4.33	5.00	1.67	3.00	.02	-.04
199.	Expensive	cheap	4.00	3.67	1.33	3.50	.38	.38*
200.	Upper	lower	5.00	2.67	1.33	1.00	.63	.32*

Note. Items 1 to 10 were "filler" items. For them, N = 1,541. For Items 11 to 200, $94 \leq N \leq 107$.
[a]These values represent the mean rating across three raters.
[b]These values represent the mean rating across two raters.
[c]The item-total correlation refers to the correlation between the item and the total score on Worksample 35 A.
*$\alpha < .01$.

correlation was somewhat lower than would be desirable for a developed test (about .30), indicating that some of the experimental items were not effective in discriminating between objective and subjective individuals. This is especially true of items that have negative item-total correlations because this indicates that subjective individuals were more likely to give the keyed responses than were objective individuals.

Item-total correlation and p-value were significantly correlated, with $r = .63$, $p < .001$. This relationship indicates that items that were more likely to elicit the keyed response (i.e., have larger p values) were more likely to be strongly correlated with Worksample 35 A scores (i.e., have larger item-total correlations).

The characteristics of an effective personality item were assessed by analyzing the relationships between the mean ratings for the items and the items' correlations with Worksample 35 A scores. As indicated earlier, items that possessed the properties of interest were expected to be more effective items and therefore have higher item-total correlations.

The left-hand column in Table 3 shows the correlations between the item property ratings and the item-total correlation. As expected, the contrast ratings were significantly correlated with the item-total correlations ($r = .58$, $p < .001$). Figure 1 shows the distribution of correlations with Worksample 35 A scores as a function of the mean contrast rating for each item. There are two things to notice in this figure. First, most of the items have contrast ratings above 3.0. As indicated earlier, this shows that overall, the experimental Personality items were generally contrasts. Second, there are few items with contrast ratings below 3.0 that showed substantial item-total correlations; most of the larger correlations with 35 A scores were associated with items that were rated above the scale's midpoint and thus can be characterized as probably or definitely contrasts.

Table 3

Correlations Between Item Properties and Item-Total Correlation

Item property	Item-total correlation[a]	Correlation with contrast partialled
Contrast (C)	.58[b]	--
Concreteness (Cr)	-.24[b]	.04
Physical reference (P)	.04	.03
Alternative response (A)	-.07	-.00
C X Cr	.48[b]	.03
C X P	.29[b]	.04
C X A	.35[b]	-.02
Cr X P	-.02	.03
Cr X A	-.16[c]	.02
P X A	.03	.03
C X Cr X P	.22[b]	.04
C X Cr X A	.27[b]	-.00
C X P X A	.21[b]	.03
Cr X P X A	-.05	.04
C X Cr X P X A	.17[c]	.03

Note. Each correlation coefficient based on $N = 191$.

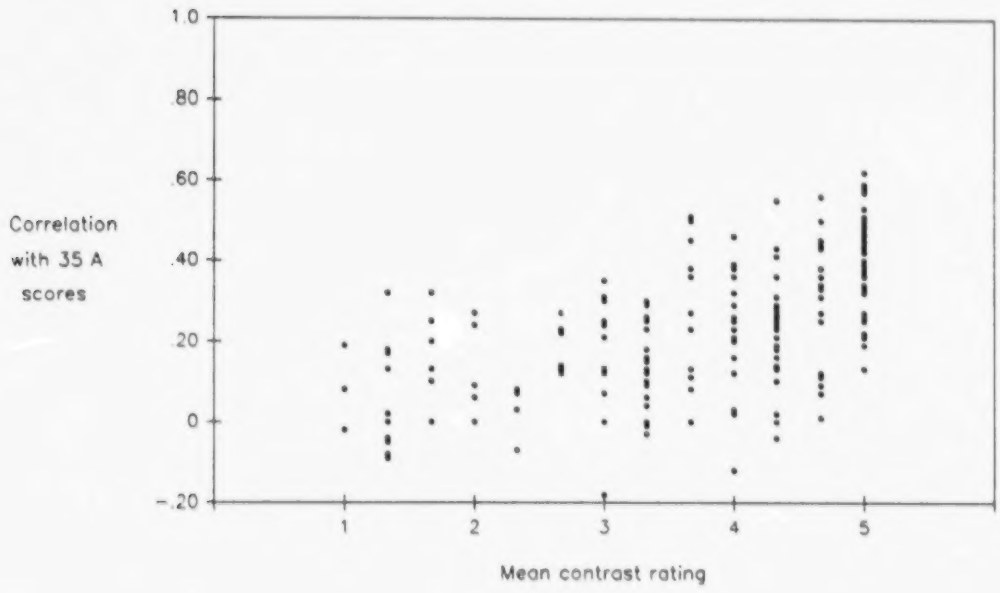
[a]The item-total correlation here refers to the correlation between the item and the total score on Worksample 35 A.

[b]$\alpha < .01$.

[c]$\alpha < .05$.

Figure 1

Distribution of Correlations With 35 A Scores as a Function of Mean Contrast Rating



Note. The correlation between mean contrast rating and item-total correlation was .58 ($p < .001$). The contrast ratings were ratings of the extent to which the word-pairs represented contrasts (i.e., opposites or coordinates; see the Introduction), with 1 = definitely not a contrast and 5 = definitely a contrast.

The concreteness ratings were also significantly correlated with item-total correlations, but in the opposite direction from the hypothesis ($r = -.24$, $p < .001$). This correlation indicates that among the items studied here, the items rated as relatively abstract (i.e., a low number on the rating scale) had relatively high item-total correlations. Similarly, items that were rated as relatively concrete (i.e., high concreteness rating) had comparatively low item-total correlations.

The mean contrast ratings were negatively correlated with the mean concreteness ratings ($r = -.45$, $p < .001$). Thus, items that were rated as definitely possessing the properties of a contrast also tended to be rated as abstract in nature. This appears to account for the negative correlation between concreteness ratings and 35 A scores (see partial correlations presented in Table 3). Note too that the mean for the concreteness scale was high (4.15), while the *SD* was low (.70), indicating that the effective range for this scale was from 3 to 5. Items rated as "may or may not be abstract" (a 3 on the scale) had relatively

high correlations with 35 A scores, and items rated as "definitely not abstract" (a 5 on the scale) had low correlations with 35 A scores. Apparently, the items rated as 3s on the concreteness scale were more likely to be contrasts than were the items rated as 5s, producing the negative correlation with the contrast property and the consequent negative correlation with 35 A scores.

To reiterate, the substantial correlation between the contrast and concreteness ratings may indicate that abstractness was a by-product of the contrast property so that relatively abstract items also tend to be more clearly contrasts than do relatively concrete items. For example, consider Items 13 and 23 shown in Table 2. Item 13 (Elbow-*knee*) has a relatively low mean contrast rating (3.33) but a high mean concreteness rating (5.00). The ratings for this item indicate that these words are definitely not abstract, nor are they clearly of a contrast nature. Item 23 (In-*out*) has a high mean contrast rating (5.00) but a low mean concreteness rating (2.33). Compared to Item 13, the ratings for Item 23 indicate that these words are relatively abstract but definitely contrast in nature. If this example reflects a stable trend in the data, then the relationship between concreteness and item-total correlation could be dissipated when the effects of the contrast property are removed from the analysis. We will return to this point when the results of the partial correlation analysis are presented (see below).

Neither the mean physical reference ratings ($r = .04$, $p > .05$) nor the mean alternative response ratings ($r = -.07$, $p > .05$) were significantly correlated with the item-total correlations. These properties appear to be unrelated to an item's effectiveness for the Personality worksample.

Although the alternative responses and physical reference properties did not show a relationship with 35 A scores, it remained a possibility that items that had these properties in combination with the contrast and concreteness properties were more effective in discriminating between objective and subjective individuals. In other words, an item that was concrete, a contrast, and had a physical referent might have been more efficient at discriminating between objective and subjective individuals than an item that was concrete, a contrast, and did *not* have a physical referent. A similar scenario could be applied to the alternative response property.

These (and other) possible interactions among item properties were examined as follows. The six two-way multiplicative interaction terms were created by multiplying the mean ratings for each pair of properties (e.g., the mean contrast rating times the mean physical reference rating). These products were then correlated with the items' correlations with 35 A scores (the item-total correlations). This procedure was repeated for the four three-way interactions (e.g., contrast X concreteness X alternative responses) and for the four-way interaction.

The left-hand column in Table 3 shows the correlations for the various interaction terms, in addition to the individual item properties. The products of the three two-way interactions involving the contrast property were all significantly correlated with the item-total correlation. The contrast by concreteness product correlated .48, $p < .001$, with item-total correlation; contrast by physical reference, .29, $p < .001$; and contrast by alternative response, .35, $p < .001$. The largest correlation that resulted from the two-way interaction analysis was between the contrast by concreteness term and correlation with 35 A scores (see Table 3). Recall that the contrast and concreteness properties were also significantly correlated with each other.

Among the remaining two-way interactions shown in Table 3 (concreteness by physical reference, concreteness by alternative responses, and physical reference by alternative responses), only the product of the concreteness by alternative responses ratings was significantly related to the item-total correlation, $r = -.16$, $p < .03$. This coefficient indicates that items that were both concrete and had high potential to generate alternative responses were not effective items. That is, these items had low correlations with scores on Worksample 35 A.

The three-way interactions that were significantly correlated with item-total correlations were contrast by concreteness by physical reference, $r = .22$, $p < .003$; contrast by concreteness by alternative response, $r = .27$, $p < .001$; and contrast by physical reference by alternative response, $r = .21$, $p < .003$. The concreteness by physical reference by alternative response interaction was not significantly correlated with item-total correlations, but the four-way interaction was significantly correlated, $r = .17$, $p < .05$ (see Table 3).

Almost all of the interaction terms that showed significant relationships with 35 A correlations involved the contrast property. The pervasive presence of the contrast property suggested that the interaction terms may have resulted in significant correlations with the item-total correlation because of the main effect of the contrast property, not because the other properties contributed to the relationship. In other words, the interaction terms were confounded with the contrast main effect.

This potential confound was examined by computing the correlations for all the predictor terms with the contrast property partialled out. The partial correlation coefficients that resulted after partialling for the contrast property are shown in the third column of Table 3. Examination of these coefficients reveals that none of the 14 variables (3 main effects, 6 two-way interactions, 4 three-way interactions, and the four-way interaction) yielded significant partial correlations with item-total correlation. The largest partial correlation was .04, for the contrast by physical reference interaction.

To summarize, the results of the subjective ratings indicate that only the contrast property resulted in a meaningful relationship with items' correlations with 35 A scores. The concreteness property and several of the interaction terms resulted in significant correlations with 35 A correlations, but the results of the partial correlation analysis indicate that these correlations resulted from a confound with the contrast property because they became nonsignificant when the contrast effect was removed. This is most evident by comparing the second and third columns of Table 3. Thus, the results of this analysis support the hypothesis that contrast items are the most efficient items for discriminating between objective and subjective individuals. They do not, however, support the notions that good items are concrete, elicit alternative responses, and have physical or sensational referents.

Construction of an Alternate Form of the Personality Worksample

The 150 experimental items that were studied in this project served as a pool from which to draw items that would be used as scored items in an alternate form of the Personality worksample. The first step in selecting items consisted of the elimination of items that were too similar to other items in the pool, or were too similar to items on Worksample 35 A. Items for which the stimulus and keyed response words were identical, but in the reverse order from other items in the pool, were also eliminated. For example, if the word pair Big-*small* was included in the pool, then Small-*big* was eliminated. For these reversed items, the stimulus-keyed response word pair that resulted in the strongest correlation with Worksample 35 A raw score was retained; the other pair was eliminated.

The next step was to eliminate items that did not have statistically significant correlations with Worksample 35 A scores. In Table 2, a separate item-total correlation is shown for each keyed response for each stimulus word. For the selection of scored items, the correlation was calculated across keyed responses so that a single correlation was obtained for each stimulus word. If a given response actually lowered the correlation with Worksample 35 A, that response was not included in this process. For some stimulus words, computing a combined-response correlation resulted in a change in the level of significance. In general a correlation of about .27 was required to reach significance at the .01 level.

Next, the items with p values of less than .15 were eliminated from the set because the keyed response(s) was not sufficiently frequent. The set of items was then divided into two groups: contrast items and noncontrast items. The analyses reported earlier indicated that, in general, contrast items measure objective-subjective personality better than noncontrast items. So, the contrast items in this set were designated as expected keyed items for the full-length alternate test.

In addition to these contrast items that we expect to score, we planned to include noncontrast items in the test so that the items would not all follow one pattern and the test design would not be transparent. The noncontrast items that had not been eliminated were included in this group. Although we labeled the noncontrast items on the alternate form as "filler" items, we will analyze the results for them and include them in the scoring key if they contribute positively to the correlation with Worksample 35 A.

The comments provided by the raters were used as the final selection criterion. Items were eliminated if the raters indicated that they might be offensive or in some cases if the stimulus word was a homonym. Since all the given responses for a particular stimulus word were documented, items were eliminated based on the latter comment only if there was evidence in the data that examinees responded to both homonymic words.

It should be noted that for a few of the items shown in Table 2, the keyed response did not achieve a significant item-total correlation, while responses other than the keyed response did result in a significant relationship between the item and total scores on 35 A. If these stimulus-response word pairs met all the criteria described above, then they were included in the alternate form.

About two-thirds of the items generated for this study failed to meet all the selection criteria required for expected scored items on the alternate form. For example, Question-*answer* (Item 70 in Table 2) was a contrast item and elicited the expected response from 71% of the examinees but did not yield a significant item-total correlation. Of those items that did not meet all the selection criteria for scored items, 11 were used as filler items for the alternate form. We expect that some of these filler items will achieve the criteria for scored items when we analyze the data from a large-sample experimental administration. For example, the item-total correlation for Spool-*thread* (Item 151 in Table 2) fell just short of significance at the .01 level, but the keyed response was given by 92% of the examinees tested on this item. Thus, this item was used as a filler item for the alternate form with the expectation that a significant item-total correlation might be achieved with the larger sample size.

Application of the selection criteria resulted in a set of 50 scored items for the alternate form.[1] One of the items, Leg-*arm* was included as an expected keyed item although we failed to obtain data regarding its correlation with Worksample 35 A because of a coding error.

---

[1]Of the 50 items, one item was included that had a significant item-total correlation for a response that was not the expected keyed response (Look-*see*). This item had an item-total correlation of .35. Since "see" was not the expected keyed response, ratings for this stimulus-response word pair were not available. Thus, this item did not contribute to the item-property analyses detailed above.

The 50 final items had contrast ratings that ranged from 2.0 to 5.0 with a mean of 4.5 (*SD* = .61), concreteness ratings that ranged from 2.7 to 5.0 with *M* = 3.9 (*SD* = .62), alternative response ratings that ranged from 1.0 to 5.0 with *M* = 3.3 (*SD* = 1.1), and physical reference ratings that ranged from 1.0 to 5.0 with *M* = 2.7 (*SD* = 1.3). Thus, the items selected for the alternate form were rated, on the average, as definitely contrasts (5 on the contrast scale), probably not abstract (4 on the nonabstractness scale), and as having average potential to evoke alternative responses (3 on the alternative response scale). The mean physical reference rating indicates that items selected for the alternate form tended to be judged as possibly (but not definitely) referring to body parts, physical actions, or physical sensations/perceptions (3 on the physical reference scale). Recall that the contrast property was the best indicator of items' effectiveness in discriminating between objective and subjective individuals. The items selected for the alternate form possess a high degree of this property, as indicated by the quite high mean for this scale (4.5).

The item-total correlations of items selected for the alternate form ranged from .30 to .62, with a mean item-total correlation of .43 (*SD* = .08). The average proportion of keyed responses for the 50 final items was .54 (*SD* = .17), with a range from .16 to .88. The final set of items, then, all had item-total correlations at or above the minimum acceptable value (.30), with the average correlation with Worksample 35 A scores well above this minimum.

The complete alternate form consists of 109 items. The first 10 items are nonscored "warm-up" items. Scored items begin with Item 11 and alternate every other item through Item 109. Filler items begin with Item 12 and were placed between each pair of expected scored items. Except as indicated above, these noncontrast filler items are not intended to be scored.

DISCUSSION

To reiterate, the analyses in this report indicated that the critical property for a Personality item is being a contrast. In other words, the keyed response(s) must be an opposite or a complement of the stimulus word. The other properties were correlated with item effectiveness (measured by correlation with 35 A scores) only insofar as they correlated with the contrast property. Thus, there was little support for the importance of concreteness, physical/sensational reference, and potential for alternative responses independent of the contrast dimension.

The results for the concreteness ratings were not consistent with Daniel's findings because items rated as relatively abstract had higher correlations with 35 A scores than did items rated as relatively concrete. The partial-correlation

analysis indicates that concreteness was not, in and of itself, a useful predictor of items' effectiveness for the Personality worksample. The effect of the concreteness property was confounded with the contrast property such that items receiving the highest contrast ratings tended to be relatively abstract items.

The hypotheses concerning physical reference and alternative responses were not supported, as a significant relationship between these properties and item effectiveness was not obtained. It appears that whether or not an item refers to a body part or a sensation is not predictive of the effectiveness of the item for discriminating between subjective and objective examinees. A similar conclusion applies to the alternative response property.

Daniel indicated that his findings regarding the semantic properties of good Personality items might account for the repeated failure to generate an alternate test that correlated acceptably with Worksample 35 A. The analyses presented here, again, suggest that the contrast property, and apparently only the contrast property, is crucial for item effectiveness. In other words, previous attempts to construct an acceptable alternate form probably did not fail because the items were too abstract, did not have physical referents, or were not capable of eliciting personalized responses (alternative responses). Since previous forms were composed largely of contrast items, it is not clear why they did not correlate sufficiently with Worksample 35 A (although correlations of about .70 [*Personality Worksample 35 A Manual*, p. 15] may reflect more favorably on the alternate forms than has previously been recognized).

Understanding the lower-than-reliability correlations may require exploring potential effects that result from administration rather than differences in the content of the two forms. Alternate forms are typically administered within a relatively short period of time following Form A. If the between-form correlation is lower than the internal-consistency reliabilities of the forms, then an administration effect, in addition to differences in the content of the two forms, is a likely cause of the reduced between-form reliability. This potential effect of administration could be examined by obtaining test-retest reliability on the same form using a short-term retest interval. If the test-retest reliability under these conditions is lower than the internal reliability, then this would imply that, to some extent, different scores will always occur on separate administrations, no matter how comparable two test forms are. As noted in the *Personality Worksample 35 A Manual* (p. 15), this could be due to a coaching/learning effect from one administration to the next or to short-term variation in mood or test strategy that affects how persons respond to the test.

Finally, the alternate form described here was constructed so that the contrast property was strongly represented among the final set of scored items, and these scored items all correlated well with total scores on Worksample 35 A. From a psychometric perspective, this alternate form possesses the characteristics expected to produce an acceptable alternate form of the Personality worksample.

REFERENCES

Daniel, M. (1977). Unpublished document, Johnson O'Connor Research Foundation, New York.

Kent, G. H., & Rosanoff, A. J. (1910). A study of association in insanity. *American Journal of Insanity, 67*, 37-96; 317-390.

O'Connor, J. (1928). *Born that way*. Baltimore, MD: Williams & Wilkins.

O'Connor, J. (1934). *Psychometrics*. Cambridge, MA: Harvard University Press.

Personality Worksample 35 A manual. (1977). M. Daniel. Chicago: Johnson O'Connor Research Foundation.

Rosenthal, R. (1982). Conducting judgment studies. In K. R. Scherer & P. Ekman (Eds.), *Handbook of methods in nonverbal behavior research* (pp. 287-361). Cambridge: Cambridge University Press.

Statistical Bulletin 1988-3. *Research proposal: Developing an alternate form for Personality*. D. H. Schroeder. Chicago: Johnson O'Connor Research Foundation.

Wells, F. L. (1911). Some properties of the free association time. *Psychological Review, 18*, 1-23.

APPENDIX A

Semantic Analysis of the Items of Worksample 35 A

by Mark Daniel

Summary: The failure to construct an alternate form of the word-association worksample that correlates highly with Worksample 35 A may result, in part, from a failure to replicate certain semantic characteristics of the best significant stimuli of that worksample. Using data from the item analysis of Wks. 35 AI performed in 1973, an attempt is made to find semantic similarities among the best-discriminating test words. There is some support for a hypothesis that stimuli relating to senses or to the body are likely to be better than other test words at discriminating between examinees who score Objective and those who score Subjective.

INTRODUCTION

The analysis and interpretation presented in this report were first done three years ago, and were based on the data arising from the 1973 item analysis of Worksample 35 AI (see Technical Report 855). The research was carried out in an attempt to identify at least some of the characteristics of "good" items on the word-association test (a "good" item being one which discriminates between people who score Objective and those who score Subjective). Ever since the introduction in the 1930s of a scoring system based on significant responses, Worksample 35 A has had a very high internal reliability. However, numerous efforts to construct an alternate form which would correlate highly with Worksample 35 A have all ended in failure, the highest correlation achieved being no higher than .75. Perhaps there are semantic characteristics of the significant test items of Worksample 35 A that are not being adequately represented in the test words of alternate forms.

The most important semantic classification system for word-association items is the set of categories described by O'Connor in *Psychometrics*. These categories of test words are:

1) Opposites, limited to adjectives and adverbs.

2) Coordinates, including no adjectives or adverbs. The stimulus and response might belong to the same class but oppose one another, as in FOLLOW-Lead, HEALTH-Sickness; or the stimulus and response might be complementary, as in MAN-Woman, SALT-Pepper, or LIGHTNING-Thunder.

3) Synonyms.

4) Classifications, in which the response gives the class to which the stimulus word belongs. Example: BUTTERFLY-Insect.

5) Associations, like coordinates, but the stimulus and response do not belong to the same type or category. Example: UMBRELLA-Rain.

6) Miscellaneous. Examples: FRUIT-Apple; SEED-Plant (is the response a noun or a verb?); COTTAGE-House (is the response a synonym or a coordinate?).

O'Connor and his associates found that test words from the first two categories, Opposites and Coordinates, tend to be best at discriminating between examinees scoring Objective and those scoring Subjective. Consequently, researchers constructing alternate word-association worksamples have included many of these "contrast" items in their forms. It is possible, however, that the emphasis placed on selecting test words which have opposites or coordinates has distracted our attention from other semantic characteristics of the good stimuli of Worksample 35 A. Specifically, recently-constructed alternate forms have been saturated with words such as "In," "Up," "Under," "Yes," "His," and "Off," which have obvious opposites or coordinates but which in other ways differ from the most effective stimulus words of Worksample 35 A. For one thing, these new stimuli are often highly abstract. For another, opposites or coordinates are so commonly given that the examinee has little opportunity *not* to give the significant response. Finally, these types of stimulus words are unlikely to conjure up a personal image, which is one of the features O'Connor felt to be important in a good test item (see *Personality Worksample 35 A Manual*, p. 2).

Semantic Analysis of the Stimulus Words of Worksample 35 A

Let us examine the most significant responses to see whether there are any noticeable characteristics of items which are efficient indicators of Objectivity.

The ten responses with the highest A/D ratios (of responses having overall frequency of 20% or more) are shown in Table 1. With the exception of the very best item, City, all of the above items are closely related to the senses or the body: Clean, Yellow, and Long relate to visual impressions; Smooth and Soft (and perhaps also Clean) are tactile characteristics; Sour relates to taste or smell; Heavy, Healthy, and Hand are closely associated with the body and basic physical perceptions. None of the ten items represents an abstract idea; all are concrete. It may be that one of the characteristics of a good significant item is that it presents a concept which is related to a simple physical sense or to the body.

Table 2 shows the next ten significant items in order of A/D ratios. At least five of these items relate to simple perceptions: Dark, Rough, High, Hard, and Square. Lightning is a phenomenon which emphasizes visual perception; Thunder is the complementary, aural phenomenon. Only one of the ten (Death) presents an abstract idea, albeit one which is highly personal.

Table 3 lists all the stimulus words of Worksample 35 AI, classified according to whether they fall into the following categories: 1) words which describe a sense perception; 2) words closely related to the body; 3) words describing emotions; 4) words signifying abstractions; and 5) words not covered by the above categories but which signify concrete objects or things.

Table 1

Ten Best Items of Worksample 35 AI

Item	Stimulus	Response	A/D
29	City	Town, Country	48.0
39	Heavy	Light	22.2
32	Sour	Sweet	17.1
65	Clean	Dirty	16.5
51	Smooth	Rough	10.4
17	Soft	Hard	10.2
93	Health	Sick(ness)	9.6
58	Yellow	Green, Red, Blue, Black, White	9.5
83	Hand	Foot	8.1
24	Long	Short	7.5

Table 2

Next Ten Best Items of Worksample 35 AI

Item	Stimulus	Response	A/D
42	Ocean	Sea	7.2
81	Death	Life	6.8
28	Dark	Light	6.3
37	Tiger	Lion	5.9
77	Rough	Smooth	5.6
34	High	Low	5.3
38	River	Stream	5.2
22	Hard	Soft	4.7
74	Lightning	Thunder	4.3
30	Square	Round	4.3

A stimulus word has a coordinate if it is one of a pair of words which together make up a set: Man-Woman, Salt-Pepper, etc. Associated words are objects or things which normally occur together, but which do not necessarily compose a set: Umbrella-Rain, Hammer-Nail, Bath-Shower, City-Town, etc. It is difficult to be certain whether a response is a coordinate, an associate, or neither, and there are bound to be disagreements over the categorizations listed above, according to how the response is understood. For example, is "Stream" a synonym for "River", or is it an associated noun? The A/D ratio figures are for the response which is categorized--that is, for the stimulus "Memory" the response "Forget" has an A/D ratio of 13.5 and is in the nature of an opposite response, but "Forget" is not the most common response to the item. If a stimulus does not have one clear response on the basis of which it is being categorized, the A/D ratio cited is the highest one for the more common responses.

The eighteen "sense" stimulus words have high A/D ratios, ranging from 3.0 to 22.2, with a median of about 6.0. The ten items associated with the body also tend to have high A/D ratios (range: 1.3 - 41.7, median about 6.3). The item "Death" might be included with the "body" items, and it has a high A/D ratio (6.8).

Table 3

Categorization of Items from Worksample 35 AI

Stimulus	A/D	Stimulus	A/D
ABSTRACTIONS HAVING OPPOSITES		*SENSES*	
Rich	2.2	Long	7.5
Follow	2.5	Black	3.3
Death	6.8	Soft	10.2
Wicked	inf.	Thick	4.1
Divide	4.2	Bitter	3.1
Memory	13.5	Dark	6.3
		Sour	17.1
		Yellow	7.4
ABSTRACTIONS NOT HAVING OPPOSITES		Rough	5.6
		Square	4.3
		Smooth	10.4
Dream	2.3	Unseen	4.9
Drift	2.0	Hard	4.7
Also	1.6	High	5.3
Wish	2.6	Heavy	22.2
Fade	2.1	Clean	16.5
Crush	2.0	Tough	14.1
Trap	3.1	Slow	3.0
Crime	3.1		
Trouble	.8	*BODY*	
Delicate	3.8		
		Age	3.1
EMOTIONS		Drink	6.8
		Hand	8.1
Bashful	1.2	Tooth	5.6
Afraid	inf.	Sleep	11.1
Joy	7.8	Bath	4.3
Cry	41.7	Ride	15.6
		Cry	41.7
		Comfort	1.3
		Whistle	5.9

Table continues

Table 3 (cont'd)

Categorization of Items from Worksample 35 AI

Stimulus	A/D	Stimulus	A/D
CONCRETE HAVING COORDINATES		*OTHER*	
Man	1.9	Blossom	1.4
Sister	3.4	Scissors	1.7
Girl	2.9	Gift	2.2
King	3.9	Cedar	1.4
Table	3.2	Railroad	1.6
Salt	3.6	Cottage	2.7
Lightning	4.3	Cap	1.6
Fruit	12.0	Tobacco	1.9
		Bed	1.9
		Eagle	1.9
CONCRETE HAVING ASSOCIATES		Captain	.6
		Whiskey	3.7
		Baker	.9
		Grocery	2.3
Hammer	1.4	Bible	1.6
Roof	1.7	Tiger	5.9
Umbrella	1.3	Ocean	7.1
Omelet	1.8	Butter	2.2
Lamp	3.3	Burglar	1.9
Stem	7.8	Thief	1.4
Seed	1.9	Hash	.9
Bath	4.3	River	5.2
Window	inf.	Sailor	1.3
Cabbage	4.0	Citizen	1.4
City	29.2	Locust	3.1
Ham	4.7	Path	3.1
		Tank	1.3
		Parlor	4.4
		Athlete	3.4
		Butterfly	1.9
		Snake	5.2
		Costume	8.3
		Music	5.2

Among the six "abstract" items having opposites, two have lower A/D ratios than all but one of the items dealing with the senses or the body. Four have quite high A/D ratios, however: Death, Wicked, Divide, and Memory. The present hypothesis regarding item types does not offer a good explanation of why the item "Divide" should have a fairly high A/D ratio (4.2) while the stimulus words "Rich" and "Follow" have low ratios. Of the ten abstract items not having opposites, the median A/D ratio is only 2.2.

The eight concrete items with coordinates (median A/D ratio 3.5) and the twelve concrete items having associated nouns (median ratio 3.7) tend to have higher A/D ratios than the concrete items having neither (median 1.6). There are few highly significant items, however, among the concrete nouns with coordinates, the best being "Fruit" (12.0) and "Lightning" (4.3). Six of the concrete items having associated nouns do not have high A/D ratios, but six do: "City" (29.2), "Stem" (7.8), "Ham" (4.7), "Bath" (4.3), "Cabbage" (4.0), and "Window" (inf.). Again, there is no good answer to the question why the item "City" should have a high A/D ratio while the stimulus word "Hammer" does not.

This initial attempt to categorize the stimulus words of Worksample 35 AI suggests that the most significant items tend to be words relating to the senses or the body. Abstract words having opposites, and nouns having either associates or coordinates, are next in order of likelihood of being significant items. Abstract or concrete words without opposites or coordinates are very unlikely to be significant.

It may be that words relating to the senses or the body are intrinsically better at measuring Objectivity/Subjectivity (as defined by this worksample) than are words relating to abstract concepts. Since we are measuring a basic personality trait, this would be a reasonable hypothesis. An analysis of the significant items of alternate forms of the word-association worksample might reveal a significant difference in the proportion of them which are body-sense related, and this could reflect on the disappointing correlations of alternate forms with 35 AI.

REFERENCE

Technical Report 855. (1977). *Construction of Personality Worksample 35 AL.* M. Daniel. New York: Johnson O'Connor Research Foundation.

APPENDIX B

Administration Instructions for Personality Worksample 35 A

(Taken from the *Scoring Guide for Group Worksamples and Standard Administration of Individual Worksamples*, pp. 64-66.)

1. Environment: Classroom/office situation with TA and examinee sitting on opposite sides of a common work surface.

2. Equipment: Scoresheet. (Hold so that the examinee cannot see answers being recorded.)

3. Procedure: The TA reads aloud the stimulus words, and records the examinee's response to each word. The stimulus words should be spoken in a consistent tone of voice, with no inflection except a downward inflection at the end of the word. It is important to maintain a constant rhythm of administration. This means that the TA will sometimes have to deliver the next stimulus word while still writing down the response to the previous item. It also means that the TA should memorize the stimulus words at the beginning of the second page so that there will be no pause between pages.

Script

This is not a timed test. I'm going to say a word. You tell me the very first word that comes to mind in response to the word I say. There are no right or wrong answers on this worksample. Just say the very first word that comes to mind.

How quickly the test should go depends on the examinee. All past revisions of the test have been aimed at getting the examinee to respond naturally and rapidly with the first word that comes to his mind. The first stimulus words are words for which there is a very common response; they are there in order to get the examinee in the habit of responding rapidly, and the TA should try to maintain nearly that pace throughout the entire test. An occasional examinee will be slow in responding to practically all the items, and there is nothing the TA can do about this. Do not attempt to "push" the examinee to respond faster by giving the next stimulus word immediately after he has made his response. This usually will not work, because the examinee realizes that he can't respond as fast as you are, and so he may block. The rhythm of stimulus words and responses should be:

STIMULUS....RESPONSE....STIMULUS....RESPONSE.

The rhythm should <u>not</u> be:

STIMULUS....RESPONSE..STIMULUS....RESPONSE.

Therefore, the test should be administered in a rhythmic fashion. If the examinee does not respond in the usual rhythm, but hesitates for 5 seconds, pass over that item and give the next stimulus word. Mark a large "X" to the left of the stimulus word that was passed over. If the examinee gives a response after you have marked the "X" but before you give the next word, circle the "X", write down the response and do not go back to it later. This still counts as a block in the final total of blocks, but the circle indicates that you should not repeat the word later. At the end of the 94 words, if you have passed over any items, say:

Now we'll try again on some words where nothing came to mind the first time. See if something comes to mind this time.

If the examinee again is unable to respond within 5 seconds, put a second "X" next to that word and go on to the next. Do not read the test words more than twice. Two "X's" by one word count as only one block.

If the examinee misunderstands the stimulus word, the response is recorded, but enclosed in parentheses. After the last test item has been given, this item is included right before the re-delivery of the skipped items. Write the examinee's second response after the first one, but do not put it in parentheses. if it is a common or significant response, circle the word instead of writing it in. If it is a significant response it should be included in the scoring just as if it had been given as the first response.

If the examinee asks that a word be repeated, simply repeat it as clearly as possible. The stimulus word should never be spelled. If the examinee says he does not know the meaning of the word, the TA writes "DNK" (for "Does Not Know") in the answer space and goes on to the next item.

If the examinee says he sees a mental image, the TA should respond, **Just give the first word that comes to mind.** If the examinee gives very long responses, the TA records as much as possible without interfering with the rhythm of the administration. If necessary, the TA writes "phr" (phrase) in the answer space.

If the examinee responds with a translation of the stimulus word into another language, or with any foreign word, ask him to limit his responses to English words.

The worksample is valid only if administered in English. If you must administer the test in a foreign language, be sure to mark "NFR" (for "Not For

Research") on the scoring sheet and the office copy bargraph.

Scoring

The total number of significant responses is the raw score.

For each item, circle the word in the "SIGNIFICANT RESPONSES" or "COMMON RESPONSES" column that corresponds exactly to the examinee's response. If the examinee's response is not listed in either column write the examinee's response in the "RESPONSE GIVEN" column.

Distinguish between singular and plural words. For example, on item 53, "chair" is a significant response, but "chairs" is not. If the response is "chairs," it should be written in the column "Response Given."

Some common and significant responses include a suffix (usually "s" to indicate the plural) in parenthesis. If the examine says the word without the suffix, circle only the word, not including the suffix. If the examinee says the word with the suffix, circle only the suffix.

On item 58, "WH" stands for "white," "BLK" stands for "black," "BL" stands for "blue" and "GR" stands for "green."

Add up the number of circled significant responses. Enter that number in the space labelled "Total Significant Responses" at the bottom of the page and in the score column of both bargraphs.

Add up the number of X's in the left hand column (but only one block per item) and enter that in the space labelled "BLOCKS" on the scoring sheet and in the space labelled (B:__) on the office copy bargraph.

JOHNSON O'CONNOR RESEARCH FOUNDATION
HUMAN ENGINEERING LABORATORY